United States Department of State
Bureau of Information Resource Management (IRM)

Open Data Plan

May 30, 2014

Table of Contents

Background

On May 9, 2013 the White House set forth an Open Data Policy via OMB Memorandum M-13-13 requiring all agencies to manage data as an Asset. The policy's goals are to increase operational efficiencies at reduced costs, improve services and increase public access to government information. For data to be open, it must be machine readable using open data standards, use open licenses, and adhere to a government-wide common core metadata standard.

Requirements

The Open Data plan describes how the Department of State continues to progress in meeting the following five core deliverables of M-13-13:

- Create and maintain an Enterprise Data Inventory (EDI)
- Create and maintain a Public Data Listing
- Create a process to engage with customers to help facilitate and prioritize data release
- Document if data cannot be released
- Clarify roles and responsibilities for promoting efficient and effective data release

Enterprise Data Inventory

The Department currently manages its inventory of information technology assets through the iMATRIX system. There are entries for approximately 360 Department systems, providing a single source for Department IT investments (applications, networks and websites). It is essential for reporting on the Department's e-Gov initiatives, enabling the Department to develop an Enterprise Architecture, and support the Assessment and Authorization process used by the Department's Information Assurance Program.

To fulfill the requirement for an inventory of all enterprise data, iMATRIX has been enhanced to include space in the system record for tracking the datasets associated with an IT investment. This was accomplished by defining a new IT asset type called DATASET. Department system owners are now able to enter information on the data assets they manage, creating an inventory of dataset's Department-wide, and thus creating State's Enterprise Data Inventory (EDI). The asset type DATASET added to iMATRIX is shown in Figure 1.

OMB Required Fields*	DoS Specific Fields
Title	Size (in MB) – to determine storage
Description	Growth % - to determine future needs
Tags (pull-down list)	FISMA Categorization
Last Update	PII Information
Publisher	Mobile enabled/ready
Contact Name	Information Area (pull-down)
Contact Email	Information Group (pull-down)
Unique Identifier (assigned by iMatrix)	
Public Access Level **	*Note: Additional data elements can be captured so that DOS can obtain valuable insights into information asset*
Data Dictionary URL+	
Download URL+	
Endpoint (Web Service) +	
Format+ [csv, xml, pdf, etc...]	Applicable only for datasets that are published to the public
Spatial+	
Temporal+	

* Metadata is based on: Project Open Data (http://project-open-data.github.io/schema/)
** Determines if additional data fields are required
\+ Additional data fields used if data is to be released to the Public

Figure 1 – DATASET Asset Type in iMatrix

As part of an ongoing process, the datasets associated with existing systems will be populated as system owners update their entries in iMatrix. System owners will be required to enter the dataset information on associated with new systems as part of their initial iMatrix system entry.

Additionally, the plan will seek support from the Application and Data Coordination Working Group (ADCWG), comprised of a broad array of data stewards from across the Department, who are working towards standardizing data so that information systems can communicate more effectively, reducing the need for ad hoc data calls. The members of the ADCWG will be approached as a resource to identify additional datasets not currently listed in the EDI.

The Enterprise Metadata Repository (EMR) is notified when new data is entered into the EDI. This provides the EMR with an opportunity to collect and store additional metadata information, such as record layout, column types, permissible values and usage; in order to support the standardization of data across the Department.

Public Data Listing

The Public Data Listing allows the public to see progress made on publishing Open Data. The list also includes metadata for those datasets not made public. In addition, the Public Data Listing populates Data.gov, so the public can search data assets generated by the U.S. government. Data.gov automatically aggregates agency-managed Public Data Listings into a centralized location, using the common core metadata standards and tagging to improve searchability. The Public Data Listing is located on the www.State.gov/data page and contained in a single JSON file. The Public Data listing will be refreshed quarterly.

Customer Engagement

Identifying and engaging with key data customers to help determine the value of federal data assets can help agencies prioritize those of highest value for quickest release. Customers will be engaged through postings on the www.State.gov/open web page, and other means as appropriate. Customers include public as well as government stakeholders. Internal customers will use blogs, e-mail and Corridor (the Department social media site) to interact with data owners directly. The Department will evaluate public and private input and reflect on how to incorporate it into their data management practices. The Department will regularly review its evolving customer feedback and public engagement strategy and develop criteria for prioritizing the opening of data assets, accounting for factors such as the quantity and quality of user demand, internal management priorities, and agency mission relevance.

Non-Releasable Data

The Open Data Policy requires agencies to develop policies and processes to ensure that only the appropriate data are publicly available. If the data owner (Data Steward) determines the data should not be made publicly available because of law, regulation, or policy or because the data are subject to privacy, confidentiality, security, trade secret, contractual, or other valid restrictions to release, it must document the determination in consultation with the Office of the Legal Advisor and the FOIA process (A/GIS/IPS). Datasets will belong to one of three following categories:

- Public: Data asset is or could be made publicly available to all without restrictions.

- Restricted Public: Data asset is available under certain use restrictions. The accessLevelComment field must be filled in with details on how one can obtain access.

- Non-Public: Data asset is not available to members of the public. This category includes data assets that are only available for internal use by the Federal government, such as by a single program, single agency, or across multiple agencies. The accessLevelComment field in the metadata must contain an explanation for the reasoning behind why these data cannot be made public for non-public datasets.

Roles and Responsibilities

The roles and responsibilities are listed for the following Open Data participants:

- System Owners – The System Owner has overall responsibility for all aspects of the information system that holds data. The registered System Owner is identified in iMatrix. The System Owner is responsible for entering all of the descriptive metadata on the system including the datasets created and maintained by the information system.
- Data Stewards – The Data Steward is the person responsible for the data entered into the information system and ensures the data entered is correct and meets quality requirements for currency and accuracy. The Data Steward makes the decision as to whether the data should be Public, Restricted Public, or Non-Public. The Data Steward prepares any documentation required to establish a dataset as Restricted Public or Non-Public.
- iMATRIX System Owner– The iMatrix system owner maintains the iMatrix system which contains, as one of its functions, the Enterprise Data Inventory.
- E-Government Program Board – Ensure IT proposals meet Department and OMB IT and E-Gov strategic principles, which includes the Open Data policy.
- ITCCB – The Information Technology Change Control Board (ITCCB) manages changes to the Department of State's global IT environment. As such, the ITCCB is responsible for ensuring new IT systems and changes to existing IT systems adhere to the Open Data policy.
- Application and Data Coordination Working Group (ADCWG) – The ADCWG has an Enterprise Data Quality Initiative that addresses the accessibility, reusability, reliability relevance and overall quality of enterprise data. The metadata entered into the EDI and the data entered into the datasets will have to follow directives associated with this initiative.
- Management Policy, Rightsizing and Innovation (M/PRI) – Reviews the Open Data Plan for consistency with the Information Sharing Environment.
- Data Management – Reviews the Open Data Plan for consistency with Department data policy. Reviews dataset format and structure for new datasets entered into the EDI.
- Chief Information Officer (CIO) – The CIO is ultimately responsible for the Department-wide implementation of all Open Data requirements.

Concept of Operation

The Data Steward, which may be the System Owner (new or existing system), will identify all key datasets that can be created and published. The Data Steward captures the core metadata information about the dataset and enters it into iMatrix. When entering the core metadata, the Data Steward consults with Legal and the FOIA process (A/GIS/IPS) about the correct categorization of the data: public, restricted public, or non-public. Legal and the FOIA process will clear on the final determination if the data will be restricted or non-public. Extended metadata, like record layout or permissible values, will be entered into the Enterprise Metadata Repository as a separate action. The iMatrix system owner, the Director of the Strategic Planning Office, will designate a user that will perform the metadata extraction process on the EDI, and subsequently process the data into a JSON file. The JSON file will be published on the www.state.gov/data page. This process will be done quarterly.
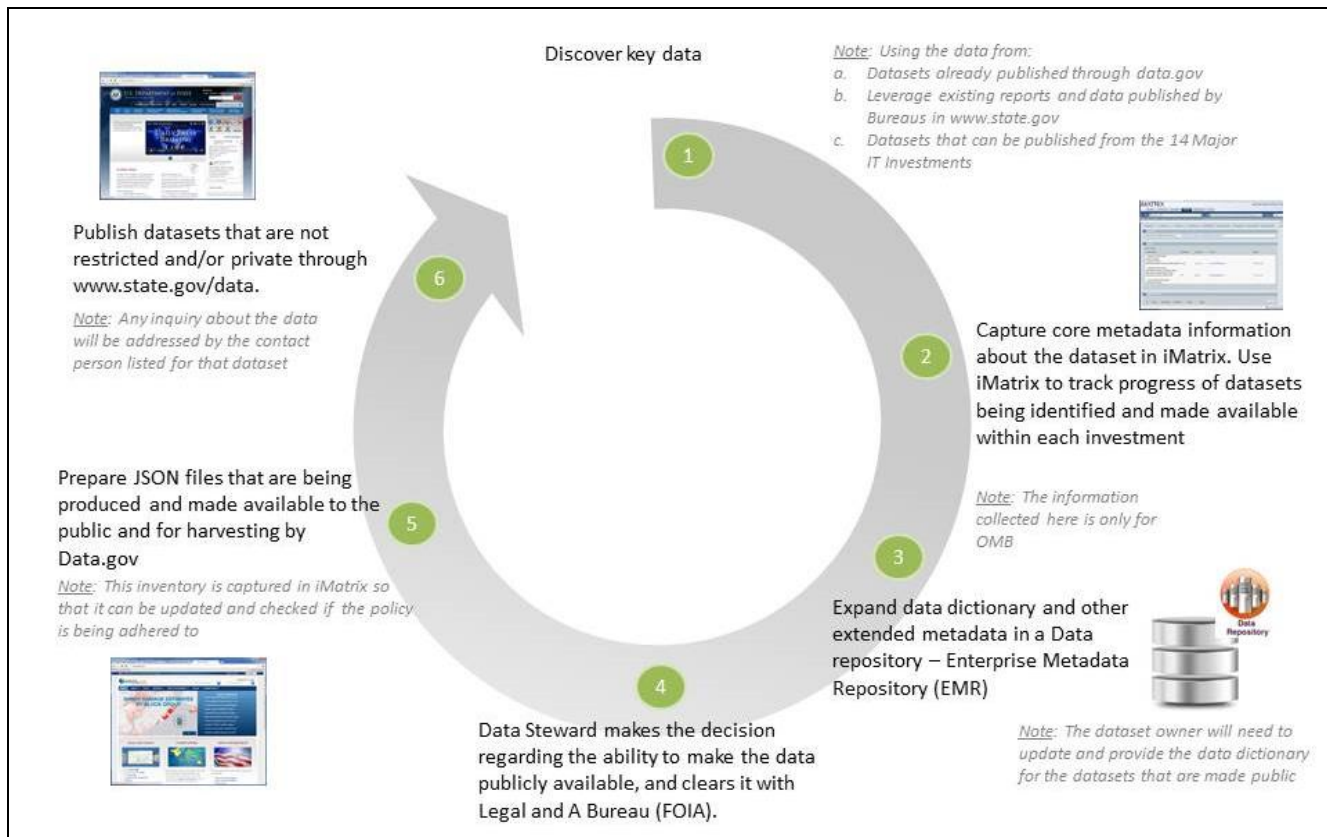
The concept of operations is shown in Figure 2.

Figure 2 – Concept of Operations

Schedule

The Department will start with the datasets owned by the bureaus/organizations in Table 1.

Owner	Notes
A/GIS/IPS/RA	Contains various information on data tagging and the policies being transmitted
ILMS	Contains various information that is used for assisting bureaus and offices in better managing the procurement
MRD	Contains some of the master reference datasets that are published for all systems within State to use
SPD	Has all of the information that has already been published through data.gov
PA	Contains the different reports and information that is published through www.state.gov
DRL	Owner of reports and data related to Human Rights
INL	Contains reports and data that has been published through their website

Table 1 – Bureaus or Offices contacted for datasets for the EDI

Every quarter the Department will target specific bureaus/offices and IT systems to make contributions to the Enterprise Data Inventory by obtaining information on datasets they are currently producing. The list of the datasets will be entered into iMATRIX. Once entered, the dataset owner is responsible for update and maintenance of the dataset and the associated metadata. Once this plan is implemented it will become part of the overall department wide Open Data management policy to support the Open Data initiative.

The schedule for the implementation of Open Data is shown in Table 2.

Milestone	Description
1	• Title: Initial Delivery • Description: The initial delivery of the Open Data Plan, the Schedule, the Enterprise Data Inventory and the Public Data Listing • Date: November 30, 2013 • Number of datasets: 113 • Open Datasets: 99
2	• Title: 1st Quarterly Update • Description: Update Open Data Plan, Schedule, Enterprise Data Inventory and Public Data Listing • Date: January 31, 2014 • Datasets Expanded: Planned – 36, Actual – 0 (113 total datasets) • Datasets Enriched: Planned – 18, Actual – 39 • Datasets Open: 0 (99 total open datasets)
3	• Title: 2nd Quarterly Update • Description: Update Open Data Plan, Schedule, Enterprise Data Inventory and Public Data Listing • Date: April 30, 2014 • Datasets Expanded: Planned – 72, Actual – 0 (113 total datasets) • Datasets Enriched: 0 • Datasets Open: 0 (99 total open datasets)
4	• Title: 3rd Quarterly Update • Description: Update Open Data Plan, Schedule, Enterprise Data Inventory and Public Data Listing • Date: July 31, 2014 • Datasets Expanded: 18 (131 total datasets) • Datasets Enriched: 18 • Datasets Open: 18 (117 total open datasets)
5	• Title: 4th Quarterly Update • Description: Update Open Data Plan, Schedule, Enterprise Data Inventory and Public Data Listing • Date: October 31, 2014 • Datasets Expanded: 36 (167 total datasets) • Datasets Enriched: 36 • Datasets Open: 18 (135 total open datasets)

Table 2 – Schedule